

Statistical Methods II: Week 12 Assignment (due 17 April, 2025)

in the following questions, PDF and CDF refer to probability density function and cumulative distribution function, respectively.

1. Suppose f and g are PDFs with respective CDFs F and G , such that there is a positive constant c ensuring $f(x) \leq cg(x)$ for all x . Suppose you generate a random variable X in the following manner.
 - (a) Generate independent samples U_1 and U_2 from the uniform distribution over the interval $(0,1)$.
 - (b) Define $X = G^{-1}(U_1)$ if $U_2 < \frac{f(U_1)}{cg(U_1)}$; otherwise repeat step (a).

What is the pdf of X ? Justify.

2. In Question 1, what is the probability that no repetition of step (a) will be needed to generate X ?
3. In Question 1, what is the probability that it will not be possible to generate a random number X even after infinite repetitions of step (a)?
4. Write an R code to simulate a sample of size $n = 100$ from the exponential distribution with mean 50, without using the function `rexp`.
5. Write an R code to simulate a sample of size $n = 100$ from the distribution with PDF f defined as $f(x) = 0.3\phi(x - 2) + 0.7\phi(x + 2)$, where ϕ is the PDF of the standard normal distribution.
6. Write an R code to simulate a sample of size $n = 100$ from the bivariate uniform distribution over the right-angled triangle with vertices at $(0,0)$, $(0,3)$ and $(4,0)$.
7. Write an R code to simulate a sample of size $n = 100$ from the bivariate uniform distribution over a circle passing through the points $(0,0)$, $(0,1)$, $(1,1)$ and $(1,0)$.
8. Write an R code to simulate a sample X_1, X_2, \dots, X_n such that $X_i - \mu$, $i = 1, 2, \dots, n$ are independent and identically distributed (iid) with the student's t distribution with 3 degrees of freedom.
9. Run the code of Question 8 for $\mu = 2$ and $n = 100$ to generate a data set and make a graphical comparison of a histogram with the actual PDF.
10. For the data of Question 9, make a graphical comparison between the empirical distribution function (EDF) and the actual CDF.
11. Compare histograms of the sample median and the sample mean generated from 1000 different samples of size 100 generated as in Question 9.
12. Compare histograms of the sample median and the sample mean generated from 1000 different samples of size 100 from the EDF of Question 10.